Trust and Manipulation in Social Networks

Manuel Förster, Ana Mauleon, Vincent Vannetelbosch

Documents de Travail du
Centre d'Economie de la Sorbonne

CES Working Papers

Trust and Manipulation in Social Networks

Manuel FÖRSTER, Ana MAULEON, Vincent VANNETELBOSCH

2013.65

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
http://centredeconomiesorbonne.univ-paris1.fr/bandeau-haut/documents-de-travail/
ISSN : 1955-611X

Trust and Manipulation in Social Networks [†]

Manuel Förster[a,b], Ana Mauleon[a,c], Vincent Vannetelbosch[a,c]

[a]CORE, Université catholique de Louvain, Belgium.
[b]CES, Université Paris 1 Panthéon-Sorbonne, France.
[c]CEREC, Université Saint-Louis – Bruxelles, Belgium.

September 18, 2013

Abstract

We investigate the role of manipulation in a model of opinion formation where agents have opinions about some common question of interest. Agents repeatedly communicate with their neighbors in the social network, can exert some effort to manipulate the trust of others, and update their opinions taking weighted averages of neighbors' opinions. The incentives to manipulate are given by the agents' preferences. We show that manipulation can modify the trust structure and lead to a connected society, and thus, make the society reaching a consensus. Manipulation fosters opinion leadership, but the manipulated agent may even gain influence on the long-run opinions. In sufficiently homophilic societies, manipulation accelerates (slows down) convergence if it decreases (increases) homophily. Finally, we investigate the tension between information aggregation and spread of misinformation. We find that if the ability of the manipulating agent is weak and the agents underselling (overselling) their information gain (lose) overall influence, then manipulation reduces misinformation and agents converge jointly to more accurate opinions about some underlying true state.

Keywords: Social networks; Trust; Manipulation; Opinion leadership; Consensus; Wisdom of crowds.

JEL Classification: D83; D85; Z13.

[†]E-mail addresses: manuel.forster@univ-paris1.fr, mauleon@fusl.ac.be, vincent.vannetelbosch@uclouvain.be. We thank Dunia Lopez-Pintado for helpful comments. Vincent Vannetelbosch and Ana Mauleon are Senior Research Associates of the National Fund for Scientific Research (FNRS). Financial support from the Doctoral Programme EDE-EM (European Doctorate in Economics - Erasmus Mundus) of the European Commission and from the Spanish Ministry of Economy and Competition under the project ECO2012-35820 are gratefully acknowledged.

1 Introduction

Individuals often rely on social connections (friends, neighbors and coworkers as well as political actors and news sources) to form beliefs or opinions on various economic, political or social issues. Every day individuals make decisions on the basis of these beliefs. For instance, when an individual goes to the polls, her choice to vote for one of the candidates is influenced by her friends and peers, her distant and close family members, and some leaders that she listens to and respects. At the same time, the support of others is crucial to enforce interests in society. In politics, majorities are needed to pass laws and in companies, decisions might be taken by a hierarchical superior. It is therefore advantageous for individuals to increase their influence on others and to *manipulate* the way others form their beliefs. This behavior is often referred to as lobbying and widely observed in society, especially in politics.[1] Hence, it is important to understand how beliefs and behaviors evolve over time when individuals can manipulate the trust of others. Can manipulation enable a segregated society to reach a consensus about some issue of broad interest? How long does it take for beliefs to reach consensus when agents can manipulate others? Can manipulation lead a society of agents who communicate and update naïvely to more efficient information aggregation?

We consider a model of opinion formation where agents repeatedly communicate with their neighbors in the social network, can exert some effort to manipulate the trust of others, and update their opinions taking weighted averages of neighbors' opinions. At each period, two agents are first selected through some deterministic manipulation sequence and can exert effort to manipulate the social trust of each other.[2] If one of them provides some costly effort to manipulate the other one, then the manipulated agent weights relatively more the belief of the agent who is manipulating her when updating her beliefs. Second, all agents communicate with their neighbors and update consequently their beliefs using the DeGroot update rule.[3] This updating process is simple. Using her (possibly manipulated) weights, an agent's new belief is the weighted average of her neighbors' beliefs (and possibly her own belief) from the previous period. When agents have no ability to manip-

[1] See Gullberg (2008) for lobbying on climate policy in the European Union, and Austen-Smith and Wright (1994) for lobbying on U.S. Supreme Court nominations.

[2] Notice that it would be possible to use more general manipulation sequences, especially probabilistic ones, since our main results in Section 4 and Section 5.1-2 will not depend on them. However, it would, at least quantitatively, affect some of the results, e.g. those on the convergence in Section 5.3.

[3] See M.H. DeGroot (1974).

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

ulate each other, the model coincides with the classical DeGroot model of opinion formation.

The DeGroot update rule assumes that agents are boundedly rational, failing to adjust correctly for repetitions and dependencies in information that they hear multiple times. Since social networks are often fairly complex, it seems reasonable to use an approach where agents fail to update beliefs correctly.[4] Chandrasekhar et al. (2012) provide evidence from a framed field experiment that DeGroot rules of thumb models best describe features of empirical social learning. They run a unique lab experiment in the field across 19 villages in rural Karnataka (India) to discriminate between the two leading classes of social learning models – Bayesian learning models versus DeGroot rules of thumb models. They find evidence that the DeGroot rule of thumb model better explains the data than the Bayesian learning model at the network level.[5] At the individual level, they find that the DeGroot rule of thumb model performs much better than Bayesian learning in explaining the actions of an individual given a history of play.[6]

Manipulation is modeled as a communicative or interactional practise, where the manipulating agent exercises some control over the manipulated agent against her will. In this sense, manipulation is illegitimate (Van Dijk, 2006). Agents only engage in manipulation if it is worth the effort. That is, agents manipulate if it brings the opinion of the society – from their point of view – sufficiently closer to their own opinion compared to the cost of manipulation. In other words, agents prefer a society holding beliefs similar to theirs, reflecting the idea that the support of others is necessary to enforce interests. We use a concrete functional form to represent these preferences that, in our view, constitutes a natural way to model lobbying incentives. However, as we will discuss subsequently, our main results will not depend on these preferences. Broadly speaking, we take lobbying activities as given in the main part of the paper and attempt to explain its consequences for society.

[4] Choi et al. (2012) report an experimental investigation of learning in three-person networks and find that already in simple three-person networks people fail to account for repeated information. They argue that the Quantal Response Equilibrium (QRE) model can account for the behavior observed in the laboratory in a variety of networks and informational settings.

[5] At the network level (i.e. when the observational unit is the sequence of actions), the Bayesian learning model explains 62% of the actions taken by individuals while the DeGroot rule of thumb model explains over 76% of the actions taken by individuals.

[6] At the individual level (i.e. when the observational unit is the action of an individual given a history), the DeGroot rule of thumb model explains nearly 87% of the actions taken by individuals given a history.

We nevertheless first analyze the decision problem of an agent having the possibility to exert effort and having preferences for manipulation as described above. We find necessary and sufficient conditions for manipulation and show that agents can have too much ability to manipulate in some situations, that is they would be better off with less ability. Second, we show that manipulation can modify the trust structure. If the society is split up into several disconnected clusters of agents and there are also some agents outside these clusters, then the latter agents might connect different clusters by manipulating the agents therein. Such an agent, previously outside any of these clusters, would not only get influential on the agents therein, but also serve as a bridge and connect them. As we show by example, this can lead to a connected society, and thus, make the society reaching a consensus.

Third, we show that manipulation fosters opinion leadership in the sense that the manipulating agent always increases her influence on the long-run beliefs. For the other agents, this is ambiguous and depends on the social network. Surprisingly, the manipulated agent may thus even gain influence on the long-run opinions. Fourth, we provide examples showing that manipulation can accelerate or slow down convergence. In particular, in sufficiently homophilic societies, i.e. societies where agents tend to trust those agents who are similar to them, and for reasonable abilities to manipulate, manipulation accelerates convergence if it decreases homophily and otherwise it slows down convergence.

Furthermore, we show that a definitive trust structure evolves in the society and in each of the disconnected clusters manipulation comes to an end and they reach a consensus (under some weak regularity condition). At some point, opinions become too similar to be manipulated. Finally, we investigate the tension between information aggregation and spread of misinformation. We find that if the ability of the manipulating agent is weak and the agents underselling their information gain and those overselling their information lose overall influence, then manipulation reduces misinformation and agents converge jointly to more accurate opinions about some underlying true state. In particular, this means that an agent that has substantial ability to manipulate can severely harm information aggregation.

Notice that our results on the trust structure,[7] on opinion leadership, on the speed of convergence and on the spread of misinformation do not depend on the preferences for manipulation. Furthermore, our result that manipulation comes to an end, eventually, only requires that agents do not manipulate if (their beliefs about) the beliefs of others are very close to their beliefs.

There is a large and growing literature on learning in social networks. Models

[7]Except for Proposition 2 (i).

of social learning either use a Bayesian perspective or exploit some plausible rule of thumb behavior.[8] We consider a model of non-Bayesian learning over a social network closely related to DeGroot (1974), DeMarzo et al. (2003), Golub and Jackson (2010) and Acemoglu et al. (2010). DeMarzo et al. (2003) consider a DeGroot rule of thumb model of opinion formation and they show that persuasion bias affects the long-run process of social opinion formation because agents fail to account for the repetition of information propagating through the network. Golub and Jackson (2010) study learning in an environment where agents receive independent noisy signals about the true state and then repeatedly communicate with each other. They find that all opinions in a large society converge to the truth if and only if the influence of the most influential agent vanishes as the society grows.[9] Acemoglu et al. (2010) investigate the tension between information aggregation and spread of misinformation. They characterize how the presence of forceful agents affects information aggregation. Forceful agents influence the beliefs of the other agents they meet, but do not change their own opinions. Under the assumption that even forceful agents obtain some information from others, they show that all beliefs converge to a stochastic consensus. They quantify the extent of misinformation by providing bounds on the gap between the consensus value and the benchmark without forceful agents where there is efficient information aggregation.[10] Friedkin (1991) studies measures to identify opinion leaders in a model related to DeGroot. Recently, Buechel et al. (2012) develop a model of opinion formation where agents may state an opinion that differs from their true opinion because agents have preferences for conformity. They find that lower conformity fosters opinion leadership. In addition, the society becomes wiser if agents who are well informed are less conform, while uninformed agents conform more with their neighbors.

We go further by allowing agents to manipulate the trust of others and we find that the implications of manipulation are non negligible for reaching a consensus, aggregating dispersed information and accelerating convergence to a consensus.

The paper is organized as follows. In Section 2 we introduce the model of opinion

[8] Acemoglu et al. (2011) develop a model of Bayesian learning over general social networks, and Acemoglu and Ozdaglar (2011) provide an overview of recent research on opinion dynamics and learning in social networks.

[9] Golub and Jackson (2012) examine how the speed of learning and best-response processes depend on homophily. They find that convergence to a consensus is slowed by the presence of homophily but is not influenced by network density.

[10] In contrast to the averaging model, Acemoglu et al. (2010) have a model of pairwise interactions. Without forceful agents, if a pair meets two periods in a row, then in the second meeting there is no information to exchange and no change in beliefs takes place.

formation. In Section 3 we analyze the decision to manipulate or not. In Section 4 we show how manipulation can change the trust in society. In Section 5 we look at the long-run effects of manipulation. In Section 6 we investigate how manipulation affects the extent of misinformation in society. In Section 7 we conclude.

2 The Model

Let $N = \{1, \ldots, n\}$ be the set of agents who have to take a decision on some issue and repeatedly communicate with their neighbors in the social network. Each agent $i \in N$ has an initial *opinion* or *belief* $x_i(0) \in \mathbb{R}$ about the issue and an initial vector of *social trust* $m_i(0) = [m_{i1}(0), \ldots, m_{in}(0)]$ with $0 \leq m_{ij}(0) \leq 1$ for all $j \in N$ and $\sum_{j \in N} m_{ij}(0) = 1$, where $m_{ij}(0)$ is the initial trust of agent i in agent j. For $i = j$, $m_{ii}(0)$ can be interpreted as how much agent i is confident in her own initial opinion. Let $\alpha_{ij} \geq 0$ denote the *ability* of agent $i \in N$ to *manipulate* agent $j \in N$, $j \neq i$.

At each period $t \in \mathbb{N}$, two agents are first selected through some deterministic manipulation sequence and can exert effort to manipulate the social trust of each other. Then, all agents communicate with their neighbors and update their beliefs. The agents' beliefs are represented by the vector $x(t) = [x_1(t), \ldots, x_n(t)]' \in \mathbb{R}^n$ and their social trust by the matrix $M(t) = [m_{ij}(t)]_{i,j \in N}$.[11] Given the vector of opinions $x(t)$ at period t, a pair of agents is picked according to some given sequence $\mathcal{E} = (E_t)_{t=0}^{\infty}$, $E_t \in \{E \in 2^N | \ |E| = 2\}$, and both agents have the possibility to manipulate each other in order to modify the trust of the other agent.[12] We write for simplicity $E_t = ij$ whenever $E_t = \{i, j\}$. We assume that each agent $i \in N$ holds some *belief about her future trust* (BT) in the next period $t + 1$, denote $\widehat{m_i}(t + 1)$, and equal to

$$\widehat{m_i}(t + 1) = m_i(t).$$

Thus, agents do not take into account the fact that the way they trust others might be manipulated. Hence, agent i's *belief about her future opinion* (BO), $\widehat{x_i}(t + 1)$, is given by

$$\widehat{x_i}(t + 1) = \widehat{m_i}(t + 1)x(t) = m_i(t)x(t).$$

At each period $t \in \mathbb{N}$, each agent $i \in E_t$ decides whether to exert effort on the other agent $j \in E_t \backslash \{i\}$ or not, $e_{ij}(t) \in \{0, 1\}$ ("no"/"yes"), according to her utility

[11] We denote the transpose of a vector (matrix) x by x'.

[12] Most of our results are robust to more general manipulation sequences, e.g. probabilistic sequences, especially our main results in Section 4 and Section 5.1-2. We mainly chose these manipulation sequences to keep the analysis in Section 2 tractable.

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

function

$$u_i^t(e_{ij}(t)) = -[x_i(t) - s_i(t+1; e_{ij}(t))]^2 - \gamma_i e_{ij}(t),$$

where $\gamma_i > 0$ is the cost for manipulating the other agent and $s_i(t+1; e_{ij}(t))$ is agent i's *belief about society's future opinion* (BS), with

$$s_i(t+1; e_{ij}(t)) = \sum_{k \in N} \widehat{m_{ik}}(t+1) x_k(t+1).$$

So, agent i evaluates her BS using her BT. That is, she views the society's opinion as a weighted average of the opinions of the agents she is trusting, and she wants her BS close to her opinion. Thus, the trade-off for agent i is between the reduction of the gap between her opinion and society's opinion due to manipulation on the one hand and the costs of manipulation γ_i on the other hand. Note that – as we will see later on – $x_j(t+1)$ depends on $e_{ij}(t)$ and that her beliefs about the future opinions of other agents are correct.[13] The decision of agent i leads to the following updated trust weights of agent j:

$$m_{jk}(t+1) = \begin{cases} m_{jk}(t) / (1 + \alpha_{ij} e_{ij}(t)) & \text{if } k \neq i \\ (m_{jk}(t) + \alpha_{ij} e_{ij}(t)) / (1 + \alpha_{ij} e_{ij}(t)) & \text{if } k = i \end{cases}.$$

If agent i manipulates agent j ($e_{ij}(t) = 1$), then j's trust in i increases according to i's ability to manipulate j (α_{ij}) and all trust weights of j are normalized. Otherwise ($e_{ij}(t) = 0$), the trust matrix does not change. Secondly, using the updated trust weights, the agents update their opinions:

$$x(t+1) = M(t+1)x(t).$$

From this equation it is clear that, given $j \in E_t$, $x_j(t+1)$ depends on $e_{ij}(t)$ since $m_j(t+1)$ does so. We can rewrite this equation as $x(t+1) = \overline{M}(t+1)x(0)$, where $\overline{M}(t+1) = M(t+1)M(t)\cdots M(1)$ denotes the *overall trust matrix*.

Remark 1. If no agent has any ability to manipulate, $\alpha_{ij} = 0$ for all $i, j \in N$, $i \neq j$, then our model reverts to the classical model of DeGroot (1974).

We now introduce the notion of *consensus*. For $G \subseteq N$, we denote by $x(t)|_G = (x_i(t))_{i \in G}$ the restriction of $x(t)$ to agents in G. Whether or not a consensus is reached in the limit depends generally on the initial opinions.

Definition 1. We say that a group of agents $G \subseteq N$ reaches a consensus given initial opinions $(x_i(0))_{i \in N}$, if there exists $x(\infty)|_G \in \mathbb{R}$ such that

$$\lim_{t \to \infty} x_i(t) = x(\infty)|_G \text{ for all } i \in G.$$

[13] Most of our results do not depend on these preferences, especially our main results in Section 4 and Section 5.1-2. They are mainly used for the analysis in Section 2 and in examples.

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

3 The Decision Problem

We study the decision problem that an agent faces when she has the opportunity to manipulate another agent. First, we derive a necessary and sufficient condition for an agent to exert effort.

Proposition 1. *Suppose that $E_t = ij$ at period t. Then, agent i manipulates agent j if and only if*

$$f_{ij}(x(t), M(t), \alpha_{ij}) := \frac{\alpha_{ij}}{(1+\alpha_{ij})^2} m_{ij}(t) \left[x_i(t) - \widehat{x_j}(t+1)\right] \cdot \left[(1+\alpha_{ij})\left[x_i(t) - s_i(t+1;0)\right] - \frac{\alpha_{ij}}{2} m_{ij}(t)\left[x_i(t) - \widehat{x_j}(t+1)\right]\right] \geq \frac{\gamma_i}{2}.$$

All proofs can be found in the appendix. Broadly speaking, the first part of f_{ij} reflects that on the one hand, an increasing difference between agent i's opinion and agent j's BO, i.e. j's future opinion in case i abstains from manipulation, fosters manipulation. On the other hand, its last part reflects that if this difference becomes too large relative to the difference between agent i's opinion and her BS given she does not manipulate (denoted by $\mathrm{BS}|_0$), then i will not manipulate. In this case, j's opinion is not important enough for i.

From Proposition 1 we can draw some intuition about the behavior of agents. Indeed, we are able to find necessary and sufficient conditions for manipulation and we provide a condition under which more ability is beneficial for the manipulator.

Corollary 1. *Suppose that $E_t = ij$ at period t.*

(*i*) *The following conditions are necessary for agent i manipulating agent j :*

- (*a*) $\alpha_{ij} > 0$,
- (*b*) $m_{ij}(t) > 0$,
- (*c*) $x_i(t) \neq \widehat{x_j}(t+1)$, *and, in particular,* $m_{ji}(t) < 1$,
- (*d*) $\mathrm{sgn}(x_i(t) - \widehat{x_j}(t+1)) = \mathrm{sgn}(x_i(t) - s_i(t+1;0))$.[14]

(*ii*) *If $\alpha_{ij}, m_{ij}(t) > 0$, then agent i manipulates agent j if*

- (*a*) $\mathrm{sgn}(x_i(t) - \widehat{x_j}(t+1)) = \mathrm{sgn}(x_i(t) - s_i(t+1;0))$ *and*

[14] For a real number $x \in \mathbb{R}$, the operator $\mathrm{sgn}(x)$ denotes the sign of x, with $\mathrm{sgn}(x) = 1$ if $x > 0$, $\mathrm{sgn}(x) = -1$ if $x < 0$ and $\mathrm{sgn}(x) = 0$ if $x = 0$.

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

(*b*)

$$|x_i(t) - s_i(t+1;0)| \geq |x_i(t) - \widehat{x_j}(t+1)| \\ \geq (1+\alpha_{ij})\sqrt{\gamma_i/[\alpha_{ij} m_{ij}(t)(2+\alpha_{ij}(2-m_{ij}(t)))]}.$$

(*iii*) *f_{ij} is strictly increasing in α_{ij} if and only if*

(*a*) $m_{ij}(t) > 0$,

(*b*) $\operatorname{sgn}(x_i(t) - \widehat{x_j}(t+1)) = \operatorname{sgn}(x_i(t) - s_i(t+1;0)) \neq 0$, *and*

(*c*) $|x_i(t) - s_i(t+1;0)| > |x_i(t) - \widehat{x_j}(t+1)| \cdot \alpha_{ij} m_{ij}(t)/(1+\alpha_{ij})$.

We now interpret Corollary 1. Part (i) first says that ability to manipulate is necessary for manipulation. Second, agent i abstains from manipulation if she does not trust agent j at all, the reason is that in this case j is not part of i's society. Third, if agent i's opinion coincides with agent j's BO, then it follows that $x_i(t) = \widehat{x_j}(t+1) = x_j(t+1)$ in case i abstains from manipulation. That is, j's future opinion and i's opinion coincide, and thus she has no incentives to manipulate. In particular, this is the case when j already trusts solely i. Fourth, agent i does not manipulate agent j if j's BO and i's $\mathrm{BS}|_0$ do not differ from i's opinion in the same direction. In other words, i does not manipulate j if her opinion lies between j's BO and i's $\mathrm{BS}|_0$. In this situation, manipulating j would even increase the gap between i's opinion and society's opinion.

Part (ii) says that when agent i has some ability to manipulate agent j and trusts her at least a bit, it is sufficient for i to manipulate j that the opinions are such that j's BO differs sufficiently from i's opinion and additionally i's $\mathrm{BS}|_0$ differs even more and into the same direction from her opinion. Hence, j's BO lies between i's opinion and i's $\mathrm{BS}|_0$.

We could expect that a higher ability to manipulate would foster manipulation. However, part (iii) shows that this is the case if and only if the necessary conditions in part (i) (apart from $\alpha_{ij} > 0$) are satisfied[15] and furthermore, j's BO differs not much more from i's opinion than i's $\mathrm{BS}|_0$. Thus, a higher ability to manipulate can hinder manipulation in situations where j's BO differs from i's opinion a lot more than i's $\mathrm{BS}|_0$. The reason is that in such situations, there is an optimal ability to manipulate that perfectly aligns i's opinion and her belief about society's future opinion given she manipulates, i.e. the first part of i's utility function vanishes.

[15] Note that $\operatorname{sgn}(x_i(t) - \widehat{x_j}(t+1)) = \operatorname{sgn}(x_i(t) - s_i(t+1;0)) \neq 0$ if and only if $x_i(t) \neq \widehat{x_j}(t+1) \wedge \operatorname{sgn}(x_i(t) - \widehat{x_j}(t+1)) = \operatorname{sgn}(x_i(t) - s_i(t+1;0))$.

Beyond this level of ability, agent i somehow manipulates too much and it leads to a worse outcome for her.

We illustrate our findings with the following example of a three-agent society.

Example 1 (Three-agent society)**.** Consider $N = \{1, 2, 3\}$ and $\gamma_i = 1/10$ for all $i \in N$. We assume that only agent 1 and 3 can manipulate: $\mathcal{E} = (13, 13, \ldots)$, where $\alpha_{13} = 3/4$ and $\alpha_{31} = 1/2$. Let $x(0) = (10, 3, 0)'$ be the vector of initial opinions and

$$M(0) = \begin{pmatrix} 3/5 & 1/5 & 1/5 \\ 1/10 & 2/5 & 1/2 \\ 0 & 3/5 & 2/5 \end{pmatrix}$$

be the initial trust matrix. First, agent 1 and 3 have the possibility to exert effort on each other. Since $f_{13}(x(0), M(0), 3/4) \approx 3.4 > 1/20 = \gamma_1/2$, agent 1 decides to do so, while agent 3 renounces since $m_{31}(0) = 0$ (see Corollary 1 (i.b)). Thus, manipulation leads to the updated trust of agent 3,

$$m_3(1) = (3/7, 12/35, 8/35)\,,$$

while the others' trust does not change,

$$M(1) = \begin{pmatrix} 3/5 & 1/5 & 1/5 \\ 1/10 & 2/5 & 1/2 \\ 3/7 & 12/35 & 8/35 \end{pmatrix}.$$

We get the following updated opinions:

$$x(1) = M(1)x(0) = (33/5, 11/5, 186/35)' \approx (6.6, 2.2, 5.3)'\,.$$

However, the classical DeGroot model gives $x_{\text{cl}}(1) = M(0)x(0) = (33/5, 11/5, 9/5)' \approx (6.6, 2.2, 1.8)'$. So, manipulation leads to a significantly different opinion of agent 3 at period 1. In addition, by Corollary 1 (iii), agent 1 always gains influence from having more ability to manipulate since

$$|x_1(0) - s_1(1;0)| = 131/25 > \frac{\alpha_{13}}{1+\alpha_{13}} \cdot 41/25 = \frac{\alpha_{13} m_{13}(0)}{1+\alpha_{13}} |x_1(0) - \widehat{x_3}(1)|$$

for all $\alpha_{13} \geq 0$. ■

4 The Trust Structure

We now investigate how manipulation can modify the structure of interaction or trust in society. We first shortly recall some graph-theoretic terminology.[16] We call

[16] See Golub and Jackson (2010).

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

a group of agents $C \subseteq N$ *minimal closed* at t if these agents only trust agents inside the group, i.e. $\sum_{j \in C} m_{ij}(t) = 1$ for all $i \in C$, and if this property does not hold for a proper subset $C' \subsetneq C$. The set of minimal closed groups at period t is denoted $\mathcal{C}(t)$ and is called the *trust structure*. A walk at period t of length $K-1$ is a sequence of agents $i_1, i_2, \ldots, i_K$ such that $m_{i_k, i_{k+1}}(t) > 0$ for all $k = 1, \ldots, K-1$. A walk is simple if $i_K \neq i_l$ for $l = 2, \ldots, K-1$, and a walk is a path if all agents are distinct. A *cycle* is a walk that starts and ends in the same agent. A cycle is *simple* if only the starting agent appears twice in the cycle. We say that a minimal closed group of agents $C \in \mathcal{C}(t)$ is *aperiodic* if the greatest common divisor[17] of the lengths of simple cycles involving agents from C is 1.[18] Note that this is fulfilled if $m_{ii}(t) > 0$ for some $i \in C$.

At each period t, we can decompose the set of agents N into minimal closed groups and agents outside these groups, the *rest of the world*, $R(t)$:

$$N = \bigcup_{C \in \mathcal{C}(t)} C \cup R(t).$$

Within minimal closed groups, all agents interact indirectly with each other, i.e. there is a path between any two agents. We say that the agents are *strongly connected*. For this reason, minimal closed groups are also called strongly connected and closed groups, see Golub and Jackson (2010). Moreover, agent $i \in N$ is part of the rest of the world $R(t)$ if and only if there is a path at period t from her to some agent in a minimal closed group $C \not\ni i$.

We say that a manipulation at period t does not change the trust structure if $\mathcal{C}(t+1) = \mathcal{C}(t)$. It also implies that $R(t+1) = R(t)$. We find that agents within a minimal closed group do only manipulate agents that are part of their group since the others are not part of "their society", a finding that clearly relies on the preferences for manipulation we are using. Contrary to this, it holds in general that only agents that are not part of a minimal closed group can change the trust structure by exerting effort. This happens if these agents exert effort on an agent within a minimal closed group. Intuitively, it means that the manipulating agent and possibly others join the group, but it might as well happen that the resulting group is not any more closed. This is the case if there is a path between the manipulating agent and some agent in another minimal closed group and it results in the group of the manipulated agent being disbanded.

[17] For a set of integers $S \subseteq \mathbb{N}$, $\gcd(S) = \max\{k \in \mathbb{N} \mid m/k \in \mathbb{N} \text{ for all } m \in S\}$ denotes the greatest common divisor.

[18] Note that if one agent in a simple cycle is from a minimal closed group, then so are all.

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

Proposition 2. *Suppose that $E_t = ij$ at period t.*

(*i*) *Let agent $i \in C \in \mathcal{C}(t)$. Then, agent $j \in C$ is a necessary condition for i manipulating j, and in this case, the trust structure does not change.*

(*ii*) *Let $i, j \in R(t)$. Then, agent i manipulating agent j does not change the trust structure.*

(*iii*) *Let $i \in R(t)$ and $j \in C \in \mathcal{C}(t)$ and suppose that there exists $C' \in \mathcal{C}(t)\backslash\{C\}$ such that there is a path from i to some $k \in C'$. Then, agent i manipulating agent j means disbanding C, i.e. $\mathcal{C}(t+1) = \mathcal{C}(t)\backslash\{C\}$.*

(*iv*) *Let $i \in R(t)$ and $j \in C \in \mathcal{C}(t)$ and suppose that there is no path from i to k for any $k \in \cup_{C' \in \mathcal{C}(t)\backslash\{C\}} C'$. Then, agent i manipulating agent j means that $R' \cup \{i\}$ joins C, i.e. $\mathcal{C}(t+1) = \mathcal{C}(t)\backslash\{C\} \cup \{C \cup R' \cup \{i\}\}$, where $R' = \{l \in R(t)\backslash\{i\} |$ there is a path from i to $l\}$.*

The following example shows that manipulation can enable a society to reach a consensus due to changes in the trust structure.

Example 2 (Consensus due to manipulation)**.** Take $N = \{1, 2, 3, 4\}$ and $\gamma_i = 1/10$ for all $i \in N$. Suppose that agent 4 meets alternately agents 1 and 3: $\mathcal{E} = (14, 34, 14, \ldots)$, with $\alpha_{41} = 1/4$, $\alpha_{43} = 1/2$ and $\alpha_{14}, \alpha_{34} > 0$. Let $x(0) = (10, 5, 5, -5)'$ be the vector of initial opinions and

$$M(0) = \begin{pmatrix} 4/5 & 1/5 & 0 & 0 \\ 2/5 & 3/5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 2/5 & 0 & 1/5 & 2/5 \end{pmatrix}$$

be the initial trust matrix. Hence, $\mathcal{C}(0) = \{\{1, 2\}, \{3\}\}$ and $R(0) = \{4\}$. At period 0, agents 1 and 4 have the possibility to exert effort on each other. By part (i) of Proposition 2, agent 1 renounces to do so. But, since $f_{41}(x(0), M(0), 1/4) \approx 11.5 > 1/20 = \gamma_4/2$, agent 4 exerts effort. This leads to the updated trust of agent 1,

$$m_1(1) = (16/25, 4/25, 0, 1/5)\,,$$

while the others' trust does not change, i.e. $m_i(1) = m_i(0)$ for $i = 2, 3, 4$, and the updated opinions become

$$x(1) = M(1)x(0) = (6.2, 7, 5, 3)'\,.$$

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

Notice that the group of agents 1 and 2 is disbanded (see part (iii) of Proposition 2). In the next period, agent 3 renounces to exert effort since she is isolated. Regarding agent 4, the conditions of part (ii) of Corollary 1 are satisfied since $s_4(2;0) \approx 5.1$ and $\widehat{x_3}(2) = 5$, i.e. she manipulates agent 3. It results in the following updated trust matrix

$$M(2) = \begin{pmatrix} 16/25 & 4/25 & 0 & 1/5 \\ 2/5 & 3/5 & 0 & 0 \\ 0 & 0 & 2/3 & 1/3 \\ 2/5 & 0 & 1/5 & 2/5 \end{pmatrix}.$$

Using part (iv) of Proposition 2, we have that N is now minimal closed, which implies that the group will reach a consensus, as we will see later on.
However, if we would start with $x(0) = (6, 5, 4, 5)'$ as initial opinions, then there would be no manipulation at all and thus, the agents would not reach a consensus. Thus, it clearly depends on the initial opinions whether or not agents reach a consensus. Notice that close opinions at the beginning – relative to the cost of manipulation – are likely to kill the incentives to exert effort. ■

5 The Long-Run Dynamics

We now look at the long-run effects of manipulation. First, we study the consequences of a single manipulation on the long-run opinion of minimal closed groups. In this context, we are interested in the role of manipulation in opinion leadership. Secondly, we investigate how manipulation affects the speed of convergence of minimal closed groups. Notice that these results will not depend on the preferences for manipulation nor on the manipulation sequence. Finally, we study the outcome of the influence process and illustrate our results by means of an example.

5.1 Opinion Leadership

Typically, an agent is called *opinion leader* if she has substantial influence on the long-run beliefs of a group. That is, if she is among the most influential agents in the group. Intuitively, manipulating others should increase her influence on the long-run beliefs and thus foster opinion leadership.

To investigate this issue, we need a measure for how directly agents trust other agents. For this purpose, we can make use of results from Markov chain theory. Let $(X_s)_{s=0}^{\infty}$ denote the homogeneous Markov chain corresponding to the transition

Matrix $M(t)$.[19] Then, the *mean first passage time* from state $i \in N$ to state $j \in N$ is defined as $\mathbb{E}[\min\{s \in \mathbb{N} \mid X_s = j\} \mid X_0 = i]$. It is the expected time the Markov chain needs to travel from i to j. In our terminology, the time to travel from i to j corresponds to the length $K-1$ of a simple walk $i = i_1, i_2, \ldots, i_K = j$ at period t from i to j, and taking the expectation means weighting the lengths of these walks with their *overall trust* $\Pi_{k=1}^{K-1} m_{i_k,i_{k+1}}(t)$. We therefore call this measure *mean simple walk length* from i to j. Intuitively, it takes small values if short simple walks have high overall trust, i.e. if agent i trusts agent j rather directly.

Definition 2. Take $i, j \in N$, $i \neq j$. The mean simple walk length at period t from agent i to agent j is given by

$$r_{ij}(t) = \mathbb{E}[\min\{s \in \mathbb{N} \mid X_s = j\} \mid X_0 = i],$$

where $(X_s)_{s=0}^{\infty}$ is the homogeneous Markov chain corresponding to $M(t)$.

Let us give some properties of the mean simple walk length.

Remark 2. Take $i, j \in N$, $i \neq j$.

(i) $r_{ij}(t) \geq 1$,

(ii) $r_{ij}(t) < +\infty$ if and only if there is a path from i to j, and, in particular, if $i, j \in C \in \mathcal{C}(t)$,

(iii) $r_{ij}(t) = 1$ if and only if $m_{ij}(t) = 1$.

Since calculating the mean simple walk length can be quite demanding using the definition, let us look at an alternative, implicit formula. Suppose that $i, j \in C \in \mathcal{C}(t)$ are two distinct agents in a minimal closed group. By part (ii) of Remark 2, the mean simple walk length is finite for all agents in that group. The simple walk from i to j with length 1 has overall trust $m_{ij}(t)$ and if it passes through k first, it has mean length $r_{kj}(t) + 1$, for $k \in C \backslash \{j\}$. Thus,

$$r_{ij}(t) = m_{ij}(t) + \sum_{k \in C \backslash \{j\}} m_{ik}(t)(r_{kj}(t) + 1).$$

Finally, applying $\sum_{k \in C} m_{ik}(t) = 1$ leads to the following result.

Lemma 1. *Take $i, j \in C \in \mathcal{C}(t)$, $i \neq j$. Then,*

$$r_{ij}(t) = 1 + \sum_{k \in C \backslash \{j\}} m_{ik}(t) r_{kj}(t).$$

[19] The agents are then interpreted as states of the Markov chain and the trust of i in j, $m_{ij}(t)$, is interpreted as the transition probability from state i to state j.

Note that computing the mean simple walk lengths using this formula amounts to solving a linear system of $|C|(|C|-1)$ equations, which has a unique solution.

We denote by $\pi(C;t)$ the probability vector of the agents' influence on the final consensus of their group $C \in \mathcal{C}(t)$ at period t, given that the group is aperiodic and the trust matrix does not change any more.[20] In this case, the group converges to

$$x(\infty)|_C = \pi(C;t)'\, x(t)|_C\,,$$

where $x(t)|_C = (x_i(t))_{i\in C}$ is the restriction of $x(t)$ to agents in C. In other words, $\pi_i(C;t)$, $i \in C$, is the influence of agent i's opinion at period t, $x_i(t)$, on the consensus of C. Each agent in a minimal closed group has at least some influence on the consensus: $\pi_i(C;t) > 0$ for all $i \in C$.[21] We now turn back to the long-run consequences of manipulation and thus, opinion leaders. We restrict our analysis to the case where both the manipulating and the manipulated agent are in the same minimal closed group. Since in this case the trust structure is preserved we can compare the influence on the long-run consensus of the group before and after manipulation.

Proposition 3. *Suppose that at period t, group $C \in \mathcal{C}(t)$ is aperiodic and agent $i \in C$ manipulates agent $j \in C$. Then, aperiodicity is preserved and the influence of agent $k \in C$ on the final consensus of her group changes as follows,*

$$\pi_k(C;t+1) - \pi_k(C;t) =$$
$$\begin{cases} [\alpha_{ij}/(1+\alpha_{ij})]\,\pi_i(C;t)\pi_j(C;t+1)\sum_{l\in C\setminus\{i\}} m_{jl}(t)r_{li}(t) & \text{if } k = i \\ [\alpha_{ij}/(1+\alpha_{ij})]\,\pi_k(C;t)\pi_j(C;t+1)\left(\sum_{l\in C\setminus\{k\}} m_{jl}(t)r_{lk}(t) - r_{ik}(t)\right) & \text{if } k \neq i \end{cases}.$$

Corollary 2. *Suppose that at period t, group $C \in \mathcal{C}(t)$ is aperiodic and agent $i \in C$ manipulates agent $j \in C$. Then,*

(*i*) *agent i always increases her long-run influence, $\pi_i(C;t+1) > \pi_i(C;t)$,*

(*ii*) *any other agent $k \neq i$ of the group can either gain or lose influence, depending on the trust matrix. She gains if and only if*

$$\sum_{l\in C\setminus\{k,i\}} m_{jl}(t)[r_{lk}(t) - r_{ik}(t)] > m_{jk}(t))r_{ik}(t),$$

(*iii*) *agent $k \neq i, j$ loses influence for sure if j trusts solely her, i.e. $m_{jk}(t) = 1$.*

[20] In the language of Markov chains, $\pi(C;t)$ is known as the unique stationary distribution of the aperiodic communication class C. Without aperiodicity, the class might fail to converge to consensus.

[21] See Golub and Jackson (2010).

Proposition 3 tells us that the change in long-run influence for any agent k depends on the ability of agent i to manipulate agent j, agent k's long-run influence now and the future influence of the manipulated agent j. When agent $k = i$, we find that this change is always positive. In this sense, manipulation fosters opinion leadership. It is large if agents (other than i) that are significantly trusted by j have a high mean simple walk length to i. To understand this better, let us recall that the long-run influence of an agent depends on how much she is trusted by agents that are trusted. Or, in other words, influential agents are influential on influential agents. Thus, there is a direct gain of influence due to an increase of trust from j and an indirect loss of influence (that is always dominated by the direct gain) due to a decrease of trust from j faced by agents that (indirectly) trust i. This explains why it is better for i if agents facing a high decrease of trust from j (those trusted much by j) do not (indirectly) trust i much (have a high mean simple walk length to i).

For any other agent $k \neq i$, it turns out that the change can be positive or negative. It is positive if, broadly speaking, j does not trust k a lot, the mean simple walk length from i to k is small and furthermore agents (other than k and i) that are significantly trusted by j have a higher mean simple walk length to k than i. In other words, it is positive if the manipulating agent, who gains influence for sure, (indirectly) trusts agent k significantly (small mean simple walk length from i to k), k does not face a high decrease of trust from j and those agents facing a high decrease from j (those trusted much by j) (indirectly) trust k less than i does.

Notice that for any agent $k \neq i, j$, this is a trade-off between an indirect gain of trust due to the increase of trust that i obtains from j, on the one hand, and an indirect loss of influence due to a decrease of trust from j faced by agents that (indirectly) trust k as well as the direct loss of influence due to a decrease of trust from j, on the other hand. In the extreme case where j only trusts k, the direct loss of influence dominates the indirect gain of influence for sure.

In particular, it means that even the manipulated agent j can gain influence. In a sense, such an agent would like to be manipulated because she trusts the "wrong" agents. For j, being manipulated is positive if agents she trusts significantly have a high mean simple walk length to her and furthermore, the mean simple walk length from i to her is small. Hence, it is positive if the manipulating agent (indirectly) trusts her significantly (small mean simple walk length from i to j) and agents facing a high decrease of trust from her (those she trusts) do not (indirectly) trust her much. Here, the trade-off is between the indirect gain of trust due to the increase of trust that i obtains from her and the indirect loss of influence due to a decrease

of trust from her faced by agents that (indirectly) trust her. Note that the gain of influence is particularly high if the manipulating agent trusts j significantly.

The next example shows that indeed in some situations an agent can gain from being manipulated in the sense that her influence on the long-run beliefs increases.

Example 3 (Being manipulated can increase influence)**.** Take $N = \{1, 2, 3\}$ with $E_0 = 13$ and $\alpha_{13} > 0$. Let

$$M(0) = \begin{pmatrix} 1/4 & 1/4 & 1/2 \\ 1/2 & 1/2 & 0 \\ 2/5 & 1/2 & 1/10 \end{pmatrix}$$

be the initial trust matrix. Notice that N is minimal closed. Suppose that agent 1 is manipulating agent 3. Then, from Proposition 3, we get

$$\begin{aligned} \pi_3(N;1) - \pi_3(N;0) &= \frac{\alpha_{13}}{1+\alpha_{13}} \pi_3(N;0)\pi_3(N;1) \sum_{l \neq 3} m_{3l}(0) r_{l3}(0) - r_{13}(0) \\ &= \frac{\alpha_{13}}{1+\alpha_{13}} \pi_3(N;0)\pi_3(N;1) \frac{7}{10} > 0, \end{aligned}$$

since $\pi_3(N;0), \pi_3(N;1) > 0$. Hence, being manipulated by agent 1 increases agent 3's influence on the long-run beliefs. The reason is that, initially, she trusts too much agent 2 – an agent that does not trust her at all. She gains influence from agent 1's increase of influence on the long-run beliefs since this agent trusts her. In other words, after being manipulated she is trusted by an agent that is trusted more. ■

5.2 Speed of Convergence

We have seen that within an aperiodic minimal closed group $C \in \mathcal{C}(t)$ agents reach a consensus given that the trust structure does not change anymore. This means that their opinions converge to a common opinion. By *speed of convergence* we mean the time that this convergence takes. That is, it is the time it takes for the expression

$$|x_i(t) - x_i(\infty)|$$

to become small. It is well known that this depends crucially on the second largest eigenvalue $\lambda_2(C;t)$ of the trust matrix $M(t)|_C$, where $M(t)|_C = (m_{ij}(t))_{i,j \in C}$ denotes the restriction of $M(t)$ to agents in C. Notice that $M(t)|_C$ is a stochastic matrix since C is minimal closed. The smaller it is in absolute value, the faster the influence

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

process converges (see Jackson, 2008). If $M(t)|_C$ is diagonalizable, then there exists a constant $K > 0$ such that[22]

$$|x_i(t+s) - x_i(\infty)| \leq K \ |\lambda_2(C;t)^s| \text{ for all } i \in C.$$

It follows from the Perron–Frobenius theorem (see Seneta, 2006) that $|\lambda_2(C;t)| < 1$.[23] Moreover, if additionally $M(t)|_C$ is nonsingular, then there exists some $i \in C$, $x_i(t)$ and $k > 0$ such that for large enough s,

$$|x_i(t+s) - x_i(\infty)| \geq k \ |\lambda_2(C;t)^s|.$$

This shows that the speed of convergence is governed by the second largest eigenvalue. To study how manipulation changes the speed of convergence, we therefore need to investigate the change in the second largest eigenvalue. However, as the above estimations indicate, this will only vaguely capture the change in speed of convergence.

For non-generic trust matrices the change in speed of convergence depends continuously on the ability to manipulate. If $M(t)|_C$ is diagonalizable and agent $i \in C$ manipulates agent $j \in C$, then it follows from the Bauer-Fike theorem (see Bauer and Fike, 1960) that there exists some $K > 0$ such that for each eigenvalue $\lambda(C;t)$ there exists an eigenvalue $\lambda(C;t+1)$ such that

$$|\lambda(C;t+1) - \lambda(C;t)| \leq K(1 - m_{ji}(t))\frac{\alpha_{ij}}{1+\alpha_{ij}}.$$

However, for a given ability, the right hand side can be large if agent j does not trust agent i a lot before manipulation since the constant K can be rather large.

Next, we want to investigate whether the second largest eigenvalue becomes larger or smaller due to manipulation. In this context, the concept of *homophily* is important, that is the tendency of people to interact relatively more with those people who are similar to them.

Definition 3. The homophily of a group of agents $G \subseteq N$ at period t is defined as

$$\text{Hom}(G;t) = \frac{1}{|G|}\left(\sum_{i,j\in G} m_{ij}(t) - \sum_{i\in G, j\notin G} m_{ij}(t)\right).$$

[22] The non-diagonalizable case is non-generic. However, a similar result holds for these matrices. See Seneta (2006).

[23] Note that $M(t)|_C$ is a non-negative primitive matrix since C is minimal closed and aperiodic.

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

The homophily of a group of agents is the normalized difference of their trust in agents inside and outside the group. Notice that a minimal closed group $C \in \mathcal{C}(t)$ attains the maximum homophily, $\text{Hom}(C;t) = 1$. As a first step, we establish the relation between manipulation and homophily. Consider a *cut of society* $(S, N\backslash S)$, $S \subseteq N, S \neq \emptyset$, into two groups of agents S and $N\backslash S$.[24] The next lemma establishes that manipulation across the cut decreases homophily, while manipulation within a group increases it.

Lemma 2. *Take a cut of society $(S, N\backslash S)$. If $i \in N$ manipulates $j \in S$ at period t, then*

(i) the homophily of S (strictly) increases if $i \in S$ (and $\sum_{k\in S} m_{jk}(t) < 1$), and

(ii) the homophily of S (strictly) decreases if $i \notin S$ (and $\sum_{k\in S} m_{jk}(t) > 0$).

Now, we come back to the speed of convergence. Given the complexity of the problem, we consider first an example of a two-agent society and show that homophily helps to explain the change in speed of convergence.

Example 4 (Speed of convergence with two agents)**.** Take $N = \{1, 2\}$ and suppose that at period t, N is minimal closed and aperiodic. Then, we have that $\lambda_2(N;t) = m_{11}(t) - m_{21}(t) = m_{22}(t) - m_{12}(t)$. Therefore, if agent i manipulates agent j at period t,

$$\begin{aligned} |\lambda_2(N;t+1)| \leq |\lambda_2(N;t)| &\Leftrightarrow |m_{11}(t+1) - m_{21}(t+1)| \leq |m_{11}(t) - m_{21}(t)| \\ &\Leftrightarrow |m_{22}(t+1) - m_{12}(t+1)| \leq |m_{22}(t) - m_{12}(t)|. \end{aligned}$$

It means that convergence is faster after manipulation if afterwards agents behave more similar, i.e. the trust both agents put on agent 1's opinion is more similar (which implies that also the trust they put on agent 2's opinion is more similar). Thus, if for instance

$$m_{22}(t) > (1 + \alpha_{12})m_{12}(t), \tag{1}$$

then agent 1 manipulating agent 2 accelerates convergence. However, if $m_{22}(t) < m_{12}(t)$, it slows down convergence since manipulation increases the already existing

[24] There exist many different notions of homophily in the literature. Our measure is similar to the one used in Golub and Jackson (2012). We can consider the average homophily $(\text{Hom}(S;t) + \text{Hom}(N\backslash S;t))/2$ with respect to the cut $(S, N\backslash S)$ as a generalization of degree-weighted homophily to general weighted averages.

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

tendency of opinions to oscillate. The more interesting case is the first one, though. We can write (1) as

$$(1+\alpha_{12})\text{Hom}(\{1\},t)+\text{Hom}(\{2\},t)>\alpha_{12},$$

that is manipulation accelerates convergence if there is sufficient aggregated homophily in the society and the ability to manipulate is not too high. The homophily of agent 1 is weighted higher since she is manipulating. ■

So, it can be that manipulation speeds up the process in the sense that a consensus is reached faster, but it can as well be the case that it is slowed down. More important, the example suggests that in a sufficiently homophilic society with reasonable abilities to manipulate, manipulation reducing homophily (i.e. across the cut, see Lemma 3) increases the speed of convergence. Notice however that manipulation increasing homophily (i.e. within one of the groups separated by the cut) is not possible in this simple setting since both groups are singletons. Therefore, let us reconsider the three-agents example to get further insights on this issue.

Example 1 (Three-agents society, continued)**.** Take $N=\{1,2,3\}$ with $\gamma_i=1/10$ for all $i\in N$, $\mathcal{E}=(13,13,\ldots)$, $\alpha_{13}=3/4$ and $\alpha_{31}=1/2$, $x(0)=(10,3,0)'$ and

$$M(0)=\begin{pmatrix}3/5 & 1/5 & 1/5\\ 1/10 & 2/5 & 1/2\\ 0 & 3/5 & 2/5\end{pmatrix}.$$

This society is homophilic, taking the cut $(\{1\},\{2,3\})$, we get

$$\text{Hom}(\{1\},0)=1/5 \text{ and } \text{Hom}(\{2,3\},0)=9/10.$$

The initial speed of convergence is $\lambda_2(N;0)=\lambda_{2,\text{cl}}=(1+\sqrt{3})/5\approx .546$. Note that under the given manipulation sequence, manipulation is across the cut and therefore it reduces homophily. We already know that only agent 1 exerts effort at period 0, which leads to

$$M(1)=\begin{pmatrix}3/5 & 1/5 & 1/5\\ 1/10 & 2/5 & 1/2\\ 3/7 & 12/35 & 8/35\end{pmatrix}$$

and $\lambda_2(N;1)=(8+\sqrt{246})/70\approx .338<\lambda_2(N;0)\approx .546$. So, convergence is faster and indeed the second group is less homophilic, $\text{Hom}(\{2,3\},1)=33/70<\text{Hom}(\{2,3\},0)=9/10$.

Let us now consider the manipulation sequence $\mathcal{E}^*=(23,23,\ldots)$ and $\alpha^*_{23},\alpha^*_{32}=1/10$, i.e. manipulation is within a group and therefore it increases homophily. At

period 0, agent 3 exerts effort since $f_{32}(x(0), M(0), 1/10) \approx .2 > 1/20 = \gamma_3/2$, while agent 2 renounces to do so since $f_{23}(x(0), M(0), 1/10) \approx .03 < 1/20 = \gamma_2/2$. This leads to

$$M^*(1) = \begin{pmatrix} 3/5 & 1/5 & 1/5 \\ 1/11 & 4/11 & 6/11 \\ 0 & 3/5 & 2/5 \end{pmatrix}$$

and $\lambda_2^*(N;1) = (10 + \sqrt{419})/55 \approx .554 > \lambda_2(N;0) \approx .546$. So, convergence is slower and indeed the second group is more homophilic, $\text{Hom}^*(\{2,3\},1) = 10/11 > \text{Hom}(\{2,3\},0) = 9/10$.[25] ■

We conclude from the examples that in sufficiently homophilic societies and for reasonable abilities to manipulate, manipulation reducing homophily increases the speed of convergence, while manipulation increasing homophily slows down convergence.

5.3 Convergence

We now determine where the process finally converges to. First, we look at the case where all agents are in the same minimal closed group and we show that manipulation comes to an end, eventually. At some point, opinions in the society become too similar to be manipulated.[26] Hence, this result holds whenever agents do not manipulate if (at least their beliefs about) the beliefs of others are very close to their own beliefs. Second, we determine the final consensus the society converges to. Let $\overline{M}(t) = I_n$ for $t < 1$, where I_n is the $n \times n$ identity matrix.

Lemma 3. *Suppose that $\mathcal{C}(0) = \{N\}$. Then, there exists $T(\mathcal{E}) \geq 0$ such that for all $t \geq T(\mathcal{E})$, $e_{ij}(t), e_{ji}(t) = 0$, where $E_t = ij$. If N is aperiodic, then the society converges to*

$$x(\infty) = \pi(N; T(\mathcal{E}))' \ \overline{M}(T(\mathcal{E}) - 1) \ x(0).$$

Notice that the outcome depends crucially on the manipulation sequence $\mathcal{E}$. Now, we turn to the general case of any trust structure. We show that after a finite number of periods, the trust structure settles down. Then, it follows from the above result

[25] Notice that these results hold for a large space of abilities to manipulate (and thus as well for the case when both agents manipulate). Only for very high abilities, manipulation can lead to oscillating opinions and thus, slow down convergence even in the case of manipulation across the cut.

[26] Without aperiodicity, the opinions might oscillate forever, but manipulation also comes to an end at some period.

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

that manipulation within the minimal closed groups that have finally been formed comes to an end. We also determine the final consensus opinion of each aperiodic minimal closed group.

Proposition 4.

(*i*) *There exists* $T(\mathcal{E}) \geq 0$ *such that for all* $t \geq T(\mathcal{E})$, $\mathcal{C}(t) = \mathcal{C}(T(\mathcal{E}))$.

(*ii*) *There exists* $\widehat{T}(\mathcal{E}) \geq T(\mathcal{E})$ *such that for all* $t \geq \widehat{T}(\mathcal{E})$, $e_{ij}(t), e_{ji}(t) = 0$ *if* $E_t = ij \subseteq C$ *for some* $C \in \mathcal{C}(T(\mathcal{E}))$. *Moreover, the agents in an aperiodic group* $C \in \mathcal{C}(T(\mathcal{E}))$ *converge to*

$$x(\infty)|_C = \pi(C; \widehat{T}(\mathcal{E}))' \; M(\widehat{T}\,(\mathcal{E}) - 1)|_C \;\; \cdots \;\; M(1)|_C \;\; x(0)|_C \,.$$

In what follows we use $T(\mathcal{E})$ and $\widehat{T}(\mathcal{E})$ in the above sense. We denote by $\overline{\pi}_i(C;t)$ the *overall influence* of agent i's initial opinion on the consensus of group C at period t given no more manipulation affecting C takes place. The overall influence is implicitly given by Proposition 4.

Corollary 3. *The overall influence of the initial opinion of agent* $i \in N$ *on the consensus of an aperiodic group* $C \in \mathcal{C}(T(\mathcal{E}))$ *is given by*

$$\overline{\pi}_i(C; \widehat{T}(\mathcal{E})) = \begin{cases} \left[\pi(C; \widehat{T}(\mathcal{E}))' \; M(\widehat{T}\,(\mathcal{E}) - 1)|_C \cdots M(1)|_C\right]_i & \text{if } i \in C \\ 0 & \text{if } i \notin C \end{cases} .$$

It turns out that an agent outside a minimal closed group that has finally formed can never have any influence on its consensus opinion. Finally, let us reconsider the three-agents example to illustrate the results of this section.

Example 1 (Three-agents society, continued)**.** Take $N = \{1,2,3\}$ with $\gamma_i = 1/10$ for all $i \in N$, $\mathcal{E} = (13, 13, \ldots)$, $\alpha_{13} = 3/4$ and $\alpha_{31} = 1/2$, $x(0) = (10, 3, 0)'$ and

$$M(0) = \begin{pmatrix} 3/5 & 1/5 & 1/5 \\ 1/10 & 2/5 & 1/2 \\ 0 & 3/5 & 2/5 \end{pmatrix}.$$

The vector of initial long-run influence – and of long-run influence in the classical model without manipulation – is $\pi(N;0) = \pi_{\text{cl}} \approx (.115, .462, .423)'$ and the initial speed of convergence is $\lambda_2(N;0) = \lambda_{2,\text{cl}} = (1+\sqrt{3})/5 \approx .546$. We already know that only agent 1 exerts effort at period 0, which leads to

$$M(1) = \begin{pmatrix} 3/5 & 1/5 & 1/5 \\ 1/10 & 2/5 & 1/2 \\ 3/7 & 12/35 & 8/35 \end{pmatrix}$$

and $x(1) = (33/5, 11/5, 186/35)' \approx (6.6, 2.2, 5.3)'$. Hence, $\pi(N;1) \approx (.4, .3, .3)'$ and $\lambda_2(N;1) = (8 + \sqrt{246})/70 \approx .338$. So, both agents 2 and 3 lose influence on the long-run beliefs and moreover, convergence is faster. At the next period, we find that again agent 1 exerts effort since $f_{13}(x(1), M(1), 3/4) \approx .2 > 1/20 = \gamma_1/2$, while agent 3 renounces to do so. To see why, note that $s_3(2;0) \approx 4.9$ and $\widehat{x_1}(2) \approx 5.46$ and thus, the necessary condition of part (i.d) of Corollary 1 is violated. Hence, we get

$$M(2) = \begin{pmatrix} 3/5 & 1/5 & 1/5 \\ 1/10 & 2/5 & 1/2 \\ 33/49 & 48/245 & 32/245 \end{pmatrix}$$

and $x(2) = (1/17150)(93688, 71981, 95526)' \approx (5.46, 4.2, 5.57)'$. It implies $\pi(N;2) \approx (.495, .249, .257)'$ and $\lambda_2(N;2) = (32 + \sqrt{4062})/490 \approx .195$. So, again both agents 2 and 3 lose influence on the long-run beliefs and convergence is even faster. The reason for the latter is that the first agent was not trusted a lot initially and so, the fact that she valued her own opinion substantially led to slow down convergence. It is easy to verify that from period 2 on, no more manipulation takes place, i.e $M(t) = M(2)$ for all $t \geq 2$. By Lemma 2, the society reaches the following consensus,

$$x(\infty) = \pi(N;2)' \ \overline{M}(1) \ x(0) = \pi(N;2)' \ M(1) \ x(0) \approx 5.18$$

and the influence of the agents' initial opinions on the consensus is

$$\overline{\pi}(N;2)' = \pi(N;2)' \ \overline{M}(1) \approx (.432, .286, .282).$$

Compared to this, the classical model gives $x_{\text{cl}}(\infty) = \pi'_{\text{cl}} x(0) \approx 2.54$, where $\pi_{\text{cl}} = \pi(N;0) \approx (.115, .462, .423)'$. Hence, our model leads to a long-run belief of society that is much closer to the initial opinion of agent 1 due to manipulation and moreover, we see that the agent not involved in manipulation, agent 2, loses more influence than the agent that was manipulated. ■

6 The Wisdom of Crowds

We now investigate how manipulation affects the extent of misinformation in society. We use an approach similar to Acemoglu et al. (2010) and assume that there is a *true state* $\mu = (1/n) \sum_{i \in N} x_i(0)$ that corresponds to the average of the initial opinions of the n agents in the society. So, each agent initially has the same information about the state. We also assume that the society forms one minimally closed and aperiodic group. Clearly, societies that are not connected fail to aggregate information. However, as in Example 2, we can observe a sequence of manipulations that

leads to a connected society and thus can be viewed as reducing the *extent of misinformation* in the society. Notice that our results in this section are qualitatively robust to changes in the preferences or the meeting sequence. Nevertheless, whether manipulation helps to aggregate information might hinge on the preferences and the meeting sequence as we show by example.

At a given period t, the *wisdom* of the society is measured by the difference between the true state and the consensus they would reach in case no more manipulation takes place:

$$\overline{\pi}(N;t)'\ x(0) - \mu = \sum_{i \in N} \left(\overline{\pi}_i(N;t) - \frac{1}{n} \right)\ x_i(0).$$

Hence, $\|\overline{\pi}(N;t) - (1/n)\mathbb{I}\|_2$ measures the extent of misinformation in the society, where $\mathbb{I} = (1,1,\ldots,1)' \in \mathbb{R}^n$ is a vector of 1s and $\|x\|_2 = \sqrt{\sum_{k \in N} |x_k|^2}$ is the standard Euclidean norm of $x \in \mathbb{R}^n$. We say that an agent i *undersells* (*oversells*) her information at period t if $\overline{\pi}_i(N;t) < 1/n$ ($\overline{\pi}_i(N;t) > 1/n$). In a sense, an agent underselling her information is, compared to her overall influence, (relatively) well informed.

Definition 4. A manipulation at period t reduces the extent of misinformation in society if

$$\|\overline{\pi}(N;t+1) - (1/n)\mathbb{I}\|_2 < \|\overline{\pi}(N;t) - (1/n)\mathbb{I}\|_2,$$

otherwise, it (weakly) increases the extent of misinformation.

The next lemma describes, given some agent manipulates another agent, the change in the overall influence of an agent from period t to period $t+1$.

Lemma 4. *Suppose that $\mathcal{C}(0) = \{N\}$ and N is aperiodic. For $k \in N$, at period t,*

$$\overline{\pi}_k(N;t+1) - \overline{\pi}_k(N;t) = \sum_{l=1}^{n} \overline{m}_{lk}(t)[\pi_l(N;t+1) - \pi_l(N;t)].$$

In case there is manipulation at period t, the overall influence of the initial opinion of an agent increases if the agents that overall trust her on average gain influence from the manipulation. Next, we provide conditions ensuring that a manipulation reduces the extent of misinformation in the society. First, the agent who is manipulating should not have too much ability to manipulate. Second, only agents underselling their information should gain overall influence. We say that $\overline{\pi}(N;t)$ is non-generic if for all $k \in N$ it holds that $\overline{\pi}_k(N;t) \neq 1/n$.

Proposition 5. *Suppose that $\mathcal{C}(0) = \{N\}$, N is aperiodic and $\overline{\pi}(N;t)$ is non-generic. Then, there exists $\overline{\alpha} > 0$ such that at period t, agent i manipulating agent j reduces the extent of misinformation if*

(*i*) $\alpha_{ij} \leq \overline{\alpha}$, *and*

(*ii*) $\sum_{l=1}^{n} \overline{m}_{lk}(t)[\pi_l(N;t+1) - \pi_l(N;t)] \geq 0$ *if and only if k undersells her information at period t.*

Intuitively, condition (ii) says that (relatively) well informed agents (those that undersell their information) should gain overall influence, while (relatively) badly informed agents (those that oversell their information) should lose overall influence. Then, this leads to a distribution of overall influence in the society that is more equal and hence reduces the extent of misinformation in the society – but only if i has not too much ability to manipulate j (condition (i)). Otherwise, manipulation makes some agents too influential, in particular the manipulating agent, and leads to a distribution of overall influence that is even more unequal than before. In other words, information aggregation can be severely harmed when some agents have substantial ability to manipulate.

We now introduce a true state of the world into Example 1. The first manipulation reduces the extent of misinformation since initially the first agent does not have much influence. However, the second manipulation increases it since agent 1 gets too influential.

Example 1 (Three-agent society, continued)**.** $N = \{1,2,3\}$ with $\gamma_i = 1/10$ for all $i \in N$, $\mathcal{E} = (13, 13, \ldots)$, $\alpha_{13} = 3/4$, $\alpha_{31} = 1/2$, $x(0) = (10,3,0)'$ and

$$M(0) = \begin{pmatrix} 3/5 & 1/5 & 1/5 \\ 1/10 & 2/5 & 1/2 \\ 0 & 3/5 & 2/5 \end{pmatrix}.$$

Hence, $\mu = (1/3)\sum_{i\in N} x_i(0) = 13/3 \approx 4.33$ is the true state. The vector of initial overall influence is $\overline{\pi}(N;0) = \pi(N;0) \approx (.115, .462, .423)'$. We already know that only agent 1 exerts effort at period 0, and it leads to

$$M(1) = \begin{pmatrix} 3/5 & 1/5 & 1/5 \\ 1/10 & 2/5 & 1/2 \\ 3/7 & 12/35 & 8/35 \end{pmatrix},$$

$x(1) \approx (6.6, 2.2, 5.3)'$ and $\overline{\pi}(N;1) = \pi(N;1) \approx (.4, .3, .3)'$. The manipulation has reduced the extent of misinformation in society since

$$\|\overline{\pi}(N;1) - (1/3)\mathbb{I}\|_2 \approx .08 < .27 \approx \|\overline{\pi}(N;0) - (1/3)\mathbb{I}\|_2.$$

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

This manipulation fulfills the conditions of Proposition 5 with threshold $\overline{\alpha} \approx 6.18$. So, even with a much higher ability agent 1 would have reduced the extent of misinformation since her initial influence was low. At the next period, we have that again only agent 1 exerts effort. We obtain $x(2) \approx (5.46, 4.2, 5.57)'$, $\pi(N;2) \approx (.495, .249, .257)'$ and $\overline{\pi}(N;2)' = \pi(N;2)'\overline{M}(1) = \pi(N;2)'M(1) \approx (.432, .286, .282)$. This manipulation has increased the extent of misinformation in society since

$$\|\overline{\pi}(N;2) - (1/3)\mathbb{I}\|_2 \approx .12 > .08 \approx \|\overline{\pi}(N;1) - (1/3)\mathbb{I}\|_2.$$

However, as there is no more manipulation from period 2 on, manipulation overall has reduced the extent of misinformation. Indeed, the agents reach the consensus $x(\infty) \approx 5.18$, which is closer to the true state $\mu \approx 4.33$ than the consensus they would have reached in the classical model of DeGroot, $x_{\text{cl}}(\infty) \approx 2.54$. Since the second manipulation has increased the extent of misinformation, the society would have been wiser if agent 1 had a higher cost (for instance, $\gamma_1 = 1/2$). Then, agent 1 would have renounced to manipulate in the second period and they would have reached the consensus $\overline{\pi}(N;1)'x(0) = 4.9$, which is closer to the true state. ■

7 Conclusion

We investigated the role of manipulation in a model of opinion formation where agents have beliefs about some question of interest and update them taking weighted averages of neighbors' opinions. Our analysis focused on the consequences of manipulation for the trust structure and long-run beliefs in the society, including learning.

We showed that manipulation can modify the trust structure and lead to a connected society, and thus, to consensus. Furthermore, we found that manipulation fosters opinion leadership in the sense that the manipulating agent always increases her influence on the long-run beliefs. And more surprisingly, this may even be the case for the manipulated agent. To obtain insides on the relation of manipulation and the speed of convergence, we provided examples and argued that in sufficiently homophilic societies and for reasonable abilities to manipulate, manipulation accelerates convergence if it decreases homophily and otherwise it slows down convergence.

Regarding learning, we were interested in the question whether manipulation is beneficial or harmful for information aggregation. We used an approach similar to Acemoglu et al. (2010) and showed that manipulation reduces the extent of misinformation in the society if the ability of the manipulating agent is weak and the agents

underselling their information gain and those overselling their information lose overall influence. Not surprisingly, agents that have substantial ability to manipulate can severely harm information aggregation. We should notice that manipulation has no bite if we use the approach of Golub and Jackson (2010). They studied large societies and showed that opinions converge to the true state if the influence of the most influential agent in the society is vanishing as the society grows. Under this condition, manipulation does not change convergence to the true state since its consequences are negligible compared to the size of the society. In large societies, information is aggregated before manipulation (and possibly a series of manipulations) can spread misinformation. The only way manipulation could have consequences for information aggregation in large societies would be to enable agents to manipulate a substantial proportion of the society instead of only one agent. However, these agents would certainly harm information aggregation.

It is important to remark that these results are robust to changes in the agents' preferences and in the manipulation sequence. In fact, they do not depend on either of them. They are driven by the way manipulation changes the social network, which reflects how manipulation is seen in the field of critical discourse analysis, see Van Dijk (2006). However, relaxing the restriction to manipulation of a single agent at a time is left for future work.

In contrast to this, our analysis of the decision problem of an agent having the possibility to exert effort clearly depends on the preferences. Apart from necessary and sufficient conditions for an agent to manipulate, we found that in some situations agents can have too much ability to manipulate, that is they would be better off with less ability.

Moreover, we showed that the trust structure of the society settles down, eventually. While this result is still qualitatively robust to changes in the preferences and the manipulation sequence since manipulation in our model can only increase connectedness, this is not the case any more for the finding that manipulation comes to an end in each of the minimal closed groups and they reach a consensus (under some weak regularity condition). This result cleary depends on the preferences for manipulation, but it is still qualitatively robust to preferences that somehow represent the idea that people do not manipulate if (their beliefs about) the beliefs of others are very close to their beliefs. However, agents would reach a different consensus if we change the preferences. For instance, we restrict agents to only manipulate other agents they trust – i.e. agents that are part of their "own" society – since they are myopic and do not anticipate the long-run effects of their decisions. However, it could be beneficial in the long-run for agent i to manipulate agent j even if agent

i does not trust agent j but j is influencing a lot all other agents. Regarding the manipulation sequence, our results would change if we generalize the sequence, e.g. to a stochastic sequence. As a result, the time when manipulation comes to an end would then be a random variable. And hence, also the consensus the society reaches would be a random variable.

We view our paper as first attempt in studying manipulation and misinformation in society. Our approach incorporated strategic considerations in a model of opinion formation à la DeGroot. We made several simplifying assumptions and derived results that apply to general societies. We plan to address some of the open issues in future work, e.g. extending manipulation to groups and allowing for more sophisticated agents.

A Appendix

Proof of Proposition 1.
First, we can rewrite i's belief about society's future opinion as

$$\begin{aligned} s_i(t+1; e_{ij}(t)) &= \sum_{k\in N} m_{ik}(t)x_k(t+1) = \sum_{k\in N} m_{ik}(t)m_k(t+1)x(t) \\ &= \sum_{k\neq j} m_{ik}(t)m_k(t+1)x(t) \\ &\quad + m_{ij}(t)\left(\sum_{l\neq i} m_{jl}(t+1)x_l(t) + m_{ji}(t+1)x_i(t)\right) \\ &= \sum_{k\neq j} m_{ik}(t)\widehat{x_k}(t+1) + \frac{m_{ij}(t)}{1+\alpha_{ij}e_{ij}(t)}\left(\widehat{x_j}(t+1) + \alpha_{ij}e_{ij}(t)x_i(t)\right), \end{aligned}$$

where the last equation follows from the definition of the updated trust weights. Hence, agent i manipulates agent j if and only if

$$\begin{aligned} & u_i^t(1) \geq u_i^t(0) \\ \Leftrightarrow & -\left[x_i(t) - \left[\sum_{k\neq j} m_{ik}(t)\widehat{x_k}(t+1) + \frac{m_{ij}(t)}{1+\alpha_{ij}}\left(\widehat{x_j}(t+1) + \alpha_{ij}x_i(t)\right)\right]\right]^2 \\ & \geq -\left[x_i(t) - \left[\sum_{k\neq j} m_{ik}(t)\widehat{x_k}(t+1) + m_{ij}(t)\widehat{x_j}(t+1)\right]\right]^2 + \gamma_i. \end{aligned}$$

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

Hence,

$$\begin{aligned}
& u_i^t(1) \geq u_i^t(0) \\
\Leftrightarrow & -\frac{\alpha_{ij}}{(1+\alpha_{ij})^2} m_{ij}(t) \Bigg[2x_i(t)(1+\alpha_{ij})(\widehat{x_j}(t+1) - x_i(t)) \\
& + 2(1+\alpha_{ij}) \left(\sum_{k \neq j} m_{ik}(t) \widehat{x_k}(t+1) \right) (x_i(t) - \widehat{x_j}(t+1)) \\
& + m_{ij}(t) \left[2\widehat{x_j}(t+1)(x_i(t) - \widehat{x_j}(t+1)) + \alpha_{ij}(x_i(t)^2 - \widehat{x_j}(t+1)^2) \right] \Bigg] \\
& \geq \gamma_i.
\end{aligned}$$

Hence,

$$\begin{aligned}
& u_i^t(1) \geq u_i^t(0) \\
\Leftrightarrow & \frac{\alpha_{ij}}{(1+\alpha_{ij})^2} m_{ij}(t) \Bigg[[x_i(t) - \widehat{x_j}(t+1)] \Bigg[2x_i(t)(1+\alpha_{ij}) \\
& - 2(1+\alpha_{ij}) \sum_{k \neq j} m_{ik}(t) \widehat{x_k}(t+1) - 2m_{ij}(t) \widehat{x_j}(t+1) \Bigg] \\
& - \left[x_i(t)^2 - \widehat{x_j}(t+1)^2 \right] m_{ij}(t) \alpha_{ij} \Bigg] \geq \gamma_i.
\end{aligned}$$

So,

$$\begin{aligned}
& u_i^t(1) \geq u_i^t(0) \\
\Leftrightarrow & \frac{\alpha_{ij}}{(1+\alpha_{ij})^2} m_{ij}(t) \left[x_i(t) - \widehat{x_j}(t+1) \right] \Big[(1+\alpha_{ij}) \left[x_i(t) - s_i(t+1;0) \right] \\
& - \frac{\alpha_{ij}}{2} m_{ij}(t) \left[x_i(t) - \widehat{x_j}(t+1) \right] \Big] \geq \frac{\gamma_i}{2},
\end{aligned}$$

which finishes the the proof. □

Proof of Corollary 1.

(i) By definition, $\alpha_{ij}, m_{ij}(t) \geq 0$. Therefore, we have $f_{ij}(x(t), M(t), \alpha_{ij}) = 0 < \gamma_i/2$ whenever one of the conditions (a)–(c) is not satisfied. For (d), suppose

that the condition does not hold. Then,

$$\begin{aligned}
&f_{ij}(x(t), M(t), \alpha_{ij}) \\
=&\frac{\alpha_{ij}}{(1+\alpha_{ij})^2} m_{ij}(t) \left[x_i(t) - \widehat{x_j}(t+1)\right] \Big[(1+\alpha_{ij}) \left[x_i(t) - s_i(t+1;0)\right] \\
&- \frac{\alpha_{ij}}{2} m_{ij}(t) \left[x_i(t) - \widehat{x_j}(t+1)\right] \Big] \\
=&\frac{\alpha_{ij}}{1+\alpha_{ij}} m_{ij}(t) \underbrace{\left[x_i(t) - \widehat{x_j}(t+1)\right] \left[x_i(t) - s_i(t+1;0)\right]}_{\leq 0} \\
&- \frac{\alpha_{ij}^2}{2(1+\alpha_{ij})^2} m_{ij}(t)^2 \underbrace{\left[x_i(t) - \widehat{x_j}(t+1)\right]^2}_{\geq 0} \\
\leq& 0 < \frac{\gamma_i}{2}.
\end{aligned}$$

which finishes this part.

(ii) By (a), we can write

$$\begin{aligned}
&f_{ij}(x(t), M(t), \alpha_{ij}) \\
=&\frac{\alpha_{ij}}{(1+\alpha_{ij})^2} m_{ij}(t) \left[x_i(t) - \widehat{x_j}(t+1)\right] \Big[(1+\alpha_{ij}) \left[x_i(t) - s_i(t+1;0)\right] \\
&- \frac{\alpha_{ij}}{2} m_{ij}(t) \left[x_i(t) - \widehat{x_j}(t+1)\right] \Big] \\
=&\frac{\alpha_{ij}}{(1+\alpha_{ij})} m_{ij}(t) |x_i(t) - \widehat{x_j}(t+1)| \underbrace{|x_i(t) - s_i(t+1;0)|}_{\geq |x_i(t) - \widehat{x_j}(t+1)| \text{ by } (b)} \\
&- \frac{\alpha_{ij}^2}{2(1+\alpha_{ij})^2} m_{ij}(t)^2 \left[x_i(t) - \widehat{x_j}(t+1)\right]^2 \\
\geq& \left[1 - \frac{\alpha_{ij} m_{ij}(t)}{2(1+\alpha_{ij})}\right] \frac{\alpha_{ij}}{(1+\alpha_{ij})} m_{ij}(t) \quad \underbrace{\left[x_i(t) - \widehat{x_j}(t+1)\right]^2}_{\geq \frac{(1+\alpha_{ij})^2 \gamma_i}{\alpha_{ij} m_{ij}(t)(2+\alpha_{ij}[2-m_{ij}(t)])} \text{ by } (b)} \\
\geq& \left[\frac{2+\alpha_{ij}[2-m_{ij}(t)]}{2(1+\alpha_{ij})}\right] \frac{(1+\alpha_{ij})\gamma_i}{2+\alpha_{ij}[2-m_{ij}(t)]} \\
=&\frac{\gamma_i}{2},
\end{aligned}$$

which finishes the second part.

(iii) ($\Leftarrow$) Suppose that conditions (a)–(c) hold. Then, taking the derivative of f_{ij}

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

with respect to α_{ij} gives

$$\begin{aligned}
&\frac{\partial f_{ij}}{\partial \alpha_{ij}}(x(t), M(t), \alpha_{ij}) \\
=&\frac{m_{ij}(t)}{(1+\alpha_{ij})^3}\Big((1+\alpha_{ij})\left[x_i(t) - \widehat{x_j}(t+1)\right]\left[x_i(t) - s_i(t+1;0)\right] \qquad \text{(A1)} \\
&- \alpha_{ij} m_{ij}(t)\left[x_i(t) - \widehat{x_j}(t+1)\right]^2 \Big) \\
\overset{(b)}{=}&\frac{m_{ij}(t)}{(1+\alpha_{ij})^3}\Big((1+\alpha_{ij})\left|x_i(t) - \widehat{x_j}(t+1)\right| \underbrace{\left|x_i(t) - s_i(t+1;0)\right|}_{> \frac{\alpha_{ij} m_{ij}(t)}{1+\alpha_{ij}} |x_i(t) - \widehat{x_j}(t+1)| \text{ by (c)}} \\
&- \alpha_{ij} m_{ij}(t)|x_i(t) - \widehat{x_j}(t+1)|^2 \Big) \\
\overset{(a)}{>}&0.
\end{aligned}$$

($\Rightarrow$) By equation (A1), $\partial f_{ij}/\partial \alpha_{ij} \leq 0$ if condition (a) or (b) does not hold. Furthermore, condition (c) is necessary as it can be seen from the above calculations, which finishes the proof. □

Proof of Proposition 2.

(i) The first part follows from Corollary 1 since $j \notin C$ implies $m_{ij}(t) = 0$. If $j \in C$, then manipulation does not change the trust structure since C is minimal closed.

(ii) Follows immediately since all minimal closed groups remain unchanged.

(iii) If agent i manipulates agent j, then $m_{ji}(t+1) > 0$ and thus, since by assumption there exists a path from i to k and C is minimal closed, there exists a path at t from l to k for all $l \in C \cup \{i\}$. Since C' is unchanged, it follows that $R(t+1) = R(t) \cup C$, i.e. $\mathcal{C}(t+1) = \mathcal{C}(t) \backslash \{C\}$.

(iv) If agent i manipulates agent j, then it follows that $\sum_{l \in C \cup \{i\}} m_{kl}(t+1) = 1$ for all $k \in C$ since C is minimal closed at t. Furthermore, since by assumption there is no path from i to k for any $k \in \cup_{C' \in \mathcal{C}(t) \backslash \{C\}} C'$ and by definition of R', $\sum_{l \in C \cup R' \cup \{i\}} m_{kl}(t+1) = 1$ for all $k \in R' \cup \{i\}$. Hence, it follows that $\sum_{l \in C \cup R' \cup \{i\}} m_{kl}(t+1) = 1$ for all $k \in C \cup R' \cup \{i\}$, i.e. $C \cup R' \cup \{i\}$ is closed. Note that, since C is minimal closed and i manipulates j, there is a path from k to l for all $k, l \in C \cup \{i\}$ at $t+1$. Then, by definition of R', there is also a path from k to l for all $k \in C \cup \{i\}$ and $l \in R'$. Moreover, since by assumption

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

there is no path from i to k for any $k \in \cup_{C' \in \mathcal{C}(t)\setminus\{C\}} C'$ and by definition of R', there exists a path from k to l for all $k \in R'$ and all $l \in C$. Combined, this implies that the same holds for all $k, l \in C \cup R' \cup \{i\}$. Hence, $C \cup R' \cup \{i\}$ is minimal closed, i.e. $\mathcal{C}(t+1) = \mathcal{C}(t)\setminus\{C\} \cup \{C \cup R' \cup \{i\}\}$. □

Proof of Proposition 3.

Suppose w.l.o.g. that $\mathcal{C}(t) = \{N\}$. First, note that aperiodicity is preserved since manipulation can only increase the number of simple cycles. We can write

$$M(t+1) = M(t) + e_j z(t)',$$

where e_j is the j-th unit vector, and

$$\begin{aligned} z_k(t) &= \begin{cases} \left(m_{ji}(t) + \alpha_{ij}\right) / \left(1 + \alpha_{ij}\right) - m_{ji}(t) & \text{if } k = i \\ \left(m_{jk}(t)\right) / \left(1 + \alpha_{ij}\right) - m_{jk}(t) & \text{if } k \neq i \end{cases} \\ &= \begin{cases} \alpha_{ij}(1 - m_{ji}(t)) / \left(1 + \alpha_{ij}\right) & \text{if } k = i \\ -\alpha_{ij} m_{jk}(t) / \left(1 + \alpha_{ij}\right) & \text{if } k \neq i \end{cases}. \end{aligned}$$

From Hunter (2005), we get

$$\begin{aligned} &\pi_k(N; t+1) - \pi_k(N; t) = -\pi_k(N; t)\pi_j(N; t+1) \sum_{l \neq k} z_l(t) r_{lk}(t) \\ &= \begin{cases} \alpha_{ij} / \left(1 + \alpha_{ij}\right) \pi_i(N; t)\pi_j(N; t+1) \sum_{l \neq i} m_{jl}(t) r_{li}(t) & \text{if } k = i \\ \alpha_{ij} / \left(1 + \alpha_{ij}\right) \pi_k(N; t)\pi_j(N; t+1) \left(\sum_{l \neq k} m_{jl}(t) r_{lk}(t) - r_{ik}(t) \right) & \text{if } k \neq i \end{cases}, \end{aligned}$$

which finishes the proof. □

Proof of Corollary 2.

We know that $\pi_k(C; t), \pi_k(C; t+1) > 0$ for all $k \in C$. By Corollary 1, we have $\alpha_{ij} > 0$ and $m_{ji}(t) < 1$. The latter implies $\sum_{l \in C\setminus\{i\}} m_{jl}(t) r_{li}(t) > 0$, which proves part (i). Part (ii) is obvious. Part (iii) follows since $m_{jk}(t) = 1$ implies $\sum_{l \in C\setminus\{k\}} m_{jl}(t) r_{lk}(t) = 0$. □

Proof of Lemma 2.

Suppose that $i \in S$. Since $\sum_{k\in S} m_{jk}(t) - \sum_{k\notin S} m_{jk}(t) \leq (<)1$, it follows that

$$\begin{aligned}
&\sum_{k\in S} m_{jk}(t) - \sum_{k\notin S} m_{jk}(t) \\
\leq (<) &\left(\sum_{k\in S} m_{jk}(t) - \sum_{k\notin S} m_{jk}(t)\right)/(1+\alpha_{ij}) + \alpha_{ij}/(1+\alpha_{ij}) \\
= &\left(\sum_{k\in S\setminus\{i\}} m_{jk}(t) - \sum_{k\notin S} m_{jk}(t)\right)/(1+\alpha_{ij}) + (m_{ji}(t)+\alpha_{ij})/(1+\alpha_{ij}) \\
= &\sum_{k\in S} m_{jk}(t+1) - \sum_{k\notin S} m_{jk}(t+1)
\end{aligned}$$

and hence $\text{Hom}(S;t+1) \geq (>)\text{Hom}(S;t)$, which finishes part (i). Part (ii) is analogous. □

Proof of Lemma 3.

Suppose that $\mathcal{C}(0) = \{N\}$. By Proposition 2, we know that $\mathcal{C}(t) = \{N\}$ for all $t \in \mathbb{N}$. To show that for any sequence $\mathcal{E}$, there exits $T(\mathcal{E}) \geq 0$ such that for all $t \geq T(\mathcal{E})$, $e_{ij}(t), e_{ji}(t) = 0$, where $E_t = ij$, it is enough to show

$$\max_{i,j\in N} |x_i(t) - x_j(t)| \to 0 \text{ for } t \to \infty$$

since this implies $f_{ij}(x(t), M(t), \alpha_{ij}) \to 0 < \gamma_i/2$ for $t \to \infty$ and all $i, j \in N$. Therefore, suppose to the contrary that $\max_{i,j\in N} |x_i(t) - x_j(t)| \geq d$ for some $d > 0$. Since $\max_{i,j\in N} |x_i(t) - x_j(t)| \geq \max_{i,j\in N} |x_i(t+1) - x_j(t+1)|$, we can choose d such that

$$\max_{i,j\in N} |x_i(t) - x_j(t)| \to d \text{ for } t \to \infty.$$

Consider $D = \{i \in N \mid |x_i(t) - x_j(t)| \to d \text{ for } t \to \infty \text{ for some } j \in N\}$. We can write $D = D_1 \cup D_2$ such that for $i \in D_1$ and $j \in D_2$: $|x_i(t) - x_j(t)| \to d$ for $t \to \infty$. Note that for $i, j \in D_k$, $|x_i(t) - x_j(t)| \to 0$ for $t \to \infty$. This implies that for all $i \in D_k$, $k = 1, 2$, either

$$\sum_{j\in D_k} m_{ij}(t) \to 1 \text{ for } t \to \infty \tag{A2}$$

or

$$\sum_{j\in D\setminus D_k} m_{ij}(t) \to 1 \text{ for } t \to \infty. \tag{A3}$$

Let us show that both possibilities lead to a contradiction. Suppose that (A2) holds. Since C is minimal closed, we can fix $i \in D_k$, $k \in \{1, 2\}$, such that $\sum_{j\in D_k} m_{ij}(t) < 1$

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

for all $t \in \mathbb{N}$. By (A2) and definition of D_k, we have for all $j \in D_k$

$$x_j(t) - \widehat{x}_i(t+1) = x_j(t) - m_i(t)x(t) \to 0 \text{ for } t \to \infty,$$

which implies $f_{ji}(x(t), M(t), \alpha_{ji}) \to 0 < \frac{\gamma_j}{2}$ for $t \to \infty$. Therefore, there exists $T \geq 0$ such that for all $t \geq T$ such that $E_t = ij$ and $j \in D_k$, $e_{ji}(t) = 0$. Hence,

$$\sum_{j \in D_k} m_{ij}(t) \leq \sum_{j \in D_k} m_{ij}(T) < 1 \text{ for all } t \geq T,$$

which is a contradiction to (A2). Similarly, (A3) leads to a contradiction by showing that $x_i(t) - s_i(t+1;0) \to 0$ for $t \to \infty$, which finishes this part.

For the second part, suppose that for all $t \geq T$, $e_{ij}(t), e_{ji}(t) = 0$, where $E_t = ij$. As already mentioned, given aperiodicity, this implies that the agents reach a consensus that can be written as

$$\begin{aligned} x(\infty) = \pi(N;T)'\, x(T) &= \pi(N;T)'\, M(T)\, x(T-1) \\ &= \pi(N;T)'\, M(T-1)\, \cdots\, M(1)\, x(0) \\ &= \pi(N;T)'\, \overline{M}(T-1)\, x(0), \end{aligned}$$

where the second equality follows from the fact that $\pi(N;T)$ is a left eigenvector of $M(T)$ corresponding to eigenvalue 1, which finishes the proof. □

Proof of Proposition 4.

Suppose that given any manipulation sequence $\mathcal{E}$, the sequence $(t_k)_{k=1}^K \subseteq \mathbb{N}$, $K \in \mathbb{N} \cup \{+\infty\}$ denotes the periods where the trust structure changes. By Proposition 2, it follows that for all $k = 1, \ldots, K$, either

(a) $1 \leq |\mathcal{C}(t_k+1)| < |\mathcal{C}(t_k)|$ and $|R(t_k+1)| > |R(t_k)|$, or

(b) $|\mathcal{C}(t_k+1)| = |\mathcal{C}(t_k)|$ and $0 \leq |R(t_k+1)| < |R(t_k)|$

holds. This implies that the maximal number of changes in the structure is bounded, i.e. $K < +\infty$. Hence, $T = t_K + 1$ is the desired threshold, which finishes part (i). Part (ii) follows from Lemma 2. The restriction to C of the matrices $M(t)$ in the computation of the consensus belief is due to the fact that $M(t)|_C$ is a stochastic matrix for all $t \geq 0$ since C is minimal closed in $\widehat{T}(\mathcal{E})$, which finishes the proof. □

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

Proof of Lemma 4.

We can write

$$\begin{aligned}\overline{\pi}_k(N;t+1) &= \sum_{l=1}^{n} \overline{m}_{lk}(t)\pi_l(N;t+1) \\ &= \sum_{l=1}^{n} \overline{m}_{lk}(t)[\pi_l(N;t+1) - \pi_l(N;t)] + \sum_{l=1}^{n} \overline{m}_{lk}(t)\pi_l(N;t) \\ &= \sum_{l=1}^{n} \overline{m}_{lk}(t)[\pi_l(N;t+1) - \pi_l(N;t)] + \underbrace{\sum_{l=1}^{n} \overline{m}_{lk}(t-1)\pi_l(N;t)}_{=\overline{\pi}_k(N;t)},\end{aligned}$$

where the last equality follows since $\pi(N;t)$ is a left eigenvector of $M(t)$. □

Proof of Proposition 5.

Let $N_* \subseteq N$ denote the set of agents that undersell their information at period t. By assumption, the agents in $N\backslash N_*$ oversell their information and additionally, $N_*, N\backslash N_* \neq \emptyset$. By Proposition 3, we have $\pi_k(N;t+1) - \pi_k(N;t) \to 0$ for $\alpha_{ij} \to 0$ and all $k \in N$ and thus by Lemma 3 we have

$$\overline{\pi}_k(N;t+1) - \overline{\pi}_k(N;t) \to 0 \text{ for } \alpha_{ij} \to 0 \text{ and all } k \in N. \tag{A4}$$

Let now $k \in N_*$, then by (ii) and Lemma 3, $\overline{\pi}_k(N;t+1) \geq \overline{\pi}_k(N;t)$. Hence, by (A4), there exists $\overline{\alpha}(k) > 0$ such that

$$1/n \geq \overline{\pi}_k(N;t+1) \geq \overline{\pi}_k(N;t) \text{ for all } \alpha_{ij} \leq \overline{\alpha}(k).$$

Analogously, for $k \in N\backslash N_*$, there exists $\overline{\alpha}(k) > 0$ such that

$$1/n \leq \overline{\pi}_k(N;t+1) < \overline{\pi}_k(N;t) \text{ for all } \alpha_{ij} \leq \overline{\alpha}(k).$$

Therefore, setting $\overline{\alpha} = \min_{k\in N} \overline{\alpha}(k)$, we get for $\alpha_{ij} \leq \overline{\alpha}$

$$\begin{aligned}\|\overline{\pi}(N;t) - \frac{1}{n}\mathbb{I}\|_2^2 &= \sum_{k\in N} |\overline{\pi}_k(N;t) - \frac{1}{n}|^2 \\ &= \sum_{k\in N_*} \underbrace{|\overline{\pi}_k(N;t) - \frac{1}{n}|^2}_{\geq |\overline{\pi}_k(N;t+1) - \frac{1}{n}|^2} + \sum_{k\in N\backslash N_*} \underbrace{|\overline{\pi}_k(N;t) - \frac{1}{n}|^2}_{> |\overline{\pi}_k(N;t+1) - \frac{1}{n}|^2} \\ &> \sum_{k\in N} |\overline{\pi}_k(N;t+1) - \frac{1}{n}|^2 \\ &= \|\overline{\pi}(N;t+1) - \frac{1}{n}\mathbb{I}\|_2^2,\end{aligned}$$

which finishes the proof. □

Documents de Travail du Centre d'Economie de la Sorbonne - 2013.65

References

[1] Acemoglu, D., Ozdaglar, A., ParandehGheibi, A.: Spread of (mis)information in social networks. Games and Economic Behavior **70**(2), 194-227 (2010)

[2] Acemoglu, D., Dahleh, M.A., Lobel, I., Ozdaglar, A.: Bayesian learning in social networks. Review of Economic Studies **78**(4), 1201-1236 (2011).

[3] Acemoglu, D., Ozdaglar, A.: Opinion dynamics and learning in social networks. Dynamic Games and Applications **1**(1), 3-49 (2011)

[4] Austen-Smith, D., Wright, J.: Counteractive lobbying. American Journal of Political Science **38**(1) 25-44 (1994)

[5] Bauer, F., Fike, C.: Norms and exclusion theorems. Numerische Mathematik **2**(1), 137-41 (1960)

[6] Buechel, B., Hellmann, T., Klößner, S.: Opinion dynamics under conformity. Institute of Mathematical Economics Working Paper 469, Bielefeld University (2012)

[7] Chandrasekhar, A., Larreguy, H., Xandri, J.: Testing models of social learning on networks: Evidence from a framed field experiment. Mimeo, Massachusetts Institute of Technology (2012)

[8] Choi, S., Gale, D., Kariv, S.: Social learning in networks: a quantal response equilibrium analysis of experimental data. Review of Economic Design **16**(2-3), 135-157 (2012)

[9] DeGroot, M.: Reaching a consensus. Journal of the American Statistical Association **69**(345), 118-21 (1974)

[10] DeMarzo, P., Vayanos, D., Zwiebel, J.: Persuasion bias, social influence, and unidimensional opinions. Quarterly Journal of Economics **118**(3), 909-68 (2003)

[11] Friedkin, N. E.: Theoretical foundations for centrality measures. American Journal of Sociology **96**(6), 1478-1504 (1991)

[12] Golub, B., Jackson, M.O.: Naïve learning in social networks and the wisdom of crowds. American Economic Journal: Microeconomics **2**(1), 112-49 (2010)

[13] Golub, B., Jackson, M.O.: How homophily affects the speed of learning and best-response dynamics. Quarterly Journal of Economics **127**(3), 1287-1338 (2012)

[14] Gullberg, A.T.: Lobbying friends and foes in climate policy: The case of business and environmental interest groups in the European Union. Energy Policy **36**(8), 2964-72 (2008)

[15] Hunter, J.: Stationary distributions and mean first passage times of perturbed Markov chains. Linear Algebra and its Applications **410**, 217-243 (2005)

[16] Jackson, M.O.: Social and Economic Networks. Princeton University Press, Princeton (2008)

[17] Seneta, E.: Non-Negative Matrices and Markov Chains (Revised Printing). Springer Series in Statistics. Springer, New York (2006)

[18] Van Dijk, T.: Discourse and manipulation. Discourse Society **17**(3), 359-83 (2006)